Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media

Patricia L. Moravec, Antino Kim, Alan R. Dennis, Randall K. Minas

To cite this article:
Patricia L. Moravec, Antino Kim, Alan R. Dennis, Randall K. Minas (2022) Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media. Information Systems Research

Published online in Articles in Advance 06 Jan 2022

. https://doi.org/10.1287/isre.2021.1090

http://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–21
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media

Patricia L. Moravec,[a] Antino Kim,[b] Alan R. Dennis,[b] Randall K. Minas[c]

[a] Information, Risk, and Operations Management Department, McCombs School of Business, University of Texas, Austin, Texas 78705;
[b] Operations and Decision Technologies Department, Kelley School of Business, Indiana University, Bloomington, Indiana 47405;
[c] Information Technology Management, Shidler College of Business, University of Hawai'i at Manoa, Honolulu, Hawaii 96822

Contact: patricia.moravec@mccombs.utexas.edu, https://orcid.org/0000-0001-5896-5444 (PLM); antino@indiana.edu, https://orcid.org/0000-0002-2886-0892 (AK); ardennis@indiana.edu, https://orcid.org/0000-0002-6439-6134 (ARD); rminas@hawaii.edu (RKM)

Received: April 7, 2020
Revised: December 14, 2020; July 17, 2021; October 1, 2021
Accepted: October 7, 2021
Published Online in Articles in Advance: January 6, 2022

https://doi.org/10.1287/isre.2021.1090

Abstract. Research has shown that *consuming ratings* influences purchase decisions in e-commerce and also has modest effects on belief in news articles on social media. However, we do not know if the act of *creating a rating* influences belief in online news stories. Unlike e-commerce settings in which ratings typically come from those who have personally used the product or service, social media users who submit their ratings for news articles typically lack firsthand knowledge of the events reported in the news, making it difficult for most users to rate news articles accurately. We propose that one key benefit of user ratings in the context of news on social media lies in triggering users who create the ratings (as opposed to *consume* the ratings) to realize that they lack this firsthand knowledge, thus inducing them to become more skeptical of articles they see. We asked 68 social media users to assess the believability of 42 social media articles and measured their cognitive activity using electroencephalography. We found that asking users to rate articles using a self-referential question induced them to think more critically—as indicated by increased activation in the medial prefrontal cortex and dorsolateral prefrontal cortex—and made them less likely to believe the articles. The effect extended to subsequent articles; after being asked to rate an article, users were less likely to believe other articles that followed it whether they were asked to rate them or not. Overall, our findings suggest that asking users to evaluate the truthfulness of articles using self-referential rating questions may not only produce rating information that could be used by others later in time, but also has an immediate benefit of inducing users to think more critically about all articles they see.

History: Emmanuelle Vaast, Senior Editor; Choon-Ling Sia, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2021.1090.

Keywords: disinformation • fake news • social media • reviews • ratings • system cognition • misinformation

Introduction

Fake news on social media rose to global attention during the 2016 U.S. presidential election (Barthel et al. 2016, Allcott and Gentzkow 2017). Reports of disinformation campaigns designed to influence election results have become commonplace as we grapple with the appropriate actions to take in response to disinformation and election meddling. More than 60% of adults get news from social media (primarily Facebook), and the proportion is increasing (Matsa and Shearer 2018), suggesting that the problem will likely worsen. Past research shows that users are poor at assessing whether a news article on social media is real or fake (Silverman 2016) because confirmation bias leads them to believe articles that align with their a priori beliefs and disbelieve those that do not (Kim and Dennis 2019, Moravec et al. 2019). As a result, more fake news articles are shared on social media than real news (Silverman 2016), and fake news spreads faster than real news (Vosoughi et al. 2018). The prevalence of fake news has shaken the public's trust in journalism and stirred up criticism of social media platforms for not taking action (Barthel et al. 2016).

In 2017, Facebook tried flagging individual articles when they were identified by experts as fake, but this proved to be ineffective and was later removed (Meixler 2017). Whereas research shows that consuming product ratings and reviews can influence purchase decisions (e.g., Hu et al. 2008, Zhu and Zhang 2010), ratings are less effective in the context of social media. Research shows that consuming ratings in the context of social media, such as fake news flags produced by others, can have little effect although seeing well-designed flags and reading ratings of the source that produced the article has some effect (Kim et al. 2019;

Moravec et al. 2019, 2020; Pennycook et al. 2020) with source ratings from experts having more effect than source ratings from users (Kim et al. 2019).

In this paper, we take a different approach to article and source rating. Past research focuses on the effects of *consuming ratings*; we focus on the effects of *creating ratings*. Facebook has proposed user ratings of news sources to help users consume more credible information, following the route of some e-commerce sites such as eBay and Yelp, which have users rate businesses (Dwoskin and Shaban 2018). Ratings such as these are the internet staple for credibility, quality, and value (Chevalier and Mayzlin 2006). When users rate a business, it is expected that they have personal firsthand knowledge; they have actually used the product or service.

This is the central problem with user ratings of news articles: how can users accurately rate news unless they have been personally involved with the event(s) reported? If you were not involved, do you actually know if it is true? Users without firsthand knowledge may aggregate knowledge from multiple news sources, but that knowledge is secondhand; they still have no direct personal knowledge of the validity of the news unlike other products and services they are asked to rate. This lack of personal firsthand knowledge combined with confirmation bias (the tendency to believe information that matches one's a priori beliefs) poses a serious challenge to user ratings.

Perhaps therein lies a potentially positive consequence of user ratings: could asking users a *self-referential question* about their firsthand personal knowledge of the events in news articles trigger them to think more carefully about whether a news article on social media is true or fake? Such a question is theoretically different from asking users for their opinion of truth (i.e., "do you think it is true?") in that the intent of the question is to invoke autobiographical memory retrieval (i.e., "do you know whether it is true from firsthand knowledge?"). Some users may give little thought when rating and simply provide the immediate assessment that comes to mind. However, a self-referential question may trigger some users to pause and realize that they actually have no firsthand way to assess the credibility of a news article—unless, of course, they have personally experienced the reported event—and this may cause them to be more skeptical of the articles they see, both those they are asked to rate as well as those they are not.

The effectiveness of this intervention does not depend on the intention of the producer of the article. Regardless of whether the producer, the reader, or the sharer is acting in "good faith"—such that they are acting to share true information rather than intentionally share false information—we need an intervention that may be effective for any user exposed to the article. Such an intervention relies on the consumer of the article thinking about the authenticity of the article and determining belief for themselves. The increase in cognition is heterogeneous, but an effective intervention applied randomly propagates through social media without concern for the intentions of producers or sharers, ideally reducing the overall set of users that see the article. As long as users are somewhat able to detect fiction with increased cognitive effort, this intervention benefits society by reducing belief of misinformation, and possibly its spread, because users are less likely to share articles they do not believe (Kim et al. 2019).

To see if asking users to rate the credibility of news articles using a self-referential question influences their assessment of those articles as well as their assessment of subsequent articles, we conducted an experiment using behavioral and neurophysiological measures (i.e., EEG). Our results show that, whereas users are still susceptible to confirmation bias, using a self-referential question nudges them to think more critically. Being asked to rate articles using a self-referential question triggered more activation in the frontal cortex regions—the medial prefrontal cortex (mPFC) and dorsolateral prefrontal cortex (DLPFC)—which indicate increased cognition associated with decision making, memory retrieval, autobiographical memory retrieval, and self-reference (Mansouri et al. 2007, 2009; Euston et al. 2012; Kim et al. 2014).

We also found support for a carryover effect; once users were asked to rate some articles, subsequent articles—even those without the self-referential rating question—triggered increased activation of these same frontal cortex regions and resulted in more critical assessment. This is in contrast to research on the consumption of fake news flags that has found that lack of a warning flag after seeing other flags causes users to believe headlines more (Pennycook et al. 2020). This can be detrimental when the article has not yet been fact-checked. Our results show that asking users a self-referential rating question triggers the opposite effect with the effect carrying over to other headlines that lack this request. Although crowdsourced ratings of news are potentially biased (Dennis et al. 2018), our results suggest that prompting users to rate news articles with self-referential questions may be an effective way to trigger users to think more critically and, as a result, potentially be less susceptible to fake news.

Prior Theory and Research

Fake news is defined as "news articles that are intentionally and verifiably false and could mislead readers" (Allcott and Gentzkow 2017, p. 213). Disinformation and fake news have long been a problem, but fake

news on social media became an important societal issue when it was reported that Russian intelligence agencies used social media to propagate fake news in an attempt to influence the 2016 presidential election in the United States (Barthel et al. 2016, Allcott and Gentzkow 2017). In response to fake news, a number of fact-checking initiatives have been launched (Graves 2016, Lowrey 2017), and fact checking is shown to influence the perceived credibility of an author or an article (Wintersieck 2017). Fact checking has traditionally been done by experts (e.g., PolitiFact); Facebook's fake news flag used expert fact checking on stories reported by many users (Mosseri 2016).

An alternative solution to expert fact checking is to have users rate the truthfulness of news articles and/or news sources. Just as eBay has users rate individual transactions that are aggregated into an overall seller rating, social media platforms could have users rate individual news articles, which would then be aggregated to provide an overall rating for the news source that created the articles. Alternately, users could be asked to rate the news source directly without considering a specific news article. These source ratings would then be applied to articles when they are first published, the same way that sellers on eBay and other e-commerce sites have their ratings applied to all new products they offer for sale. Research shows that source ratings influence the extent to which users believe social media stories (Kim and Dennis 2019) and influence online news consumption (Cerf 2016) although source ratings from experts have a stronger effect than source ratings from users (Kim et al. 2019).

Prior research on fact checking and news source ratings focuses on the consumption of fact checking and ratings (e.g., whether social media users' beliefs are influenced by fact checking or ratings; Wintersieck 2017). One unanswered question is whether—and how—asking a user to evaluate articles affects the user doing the assessment. Does the act of creating a rating induce greater skepticism? Does it influence users to invest more cognitive effort in assessing them? We begin by considering how users process information on social media and the cognitive process that takes place when users are prompted to evaluate articles.

Information Processing on Social Media

Some individuals use social media for goal-driven utilitarian reasons (e.g., connecting with potential employers on LinkedIn), but most individuals use social media such as Facebook and Twitter for hedonic purposes, such as seeking entertainment or connecting with friends (Sledgianowski and Kulviwat 2009, Johnson and Kaye 2015, Panger 2018). Individuals in a hedonic mindset are less likely to consider information critically than those in a utilitarian mindset (Hirschman and Holbrook 1982, Cotte et al. 2006). Research suggests that users "deactivate" on social media and tend not to exert cognitive effort as they use social media for hedonic purposes (Sledgianowski and Kulviwat 2009, Johnson and Kaye 2015, Panger 2018). Simply put, users do not engage deeply with what they see on social media. A study of Twitter showed that 59% of article retweets (the primary form of sharing on Twitter) were done without the user clicking on the article link (Gabielkov et al. 2016); that is, 59% of the time when someone shared an article, the person did so *without* reading it.

On Facebook, users are more likely to see articles that align with their existing beliefs (The Wall Street Journal 2016), thanks to Facebook's algorithms that are designed to identify and display content that matches users' preferences in order to maximize their enjoyment and time online. Such a process causes a decrease in the range of information that Facebook users encounter, and as a result, they often exist in information "bubbles"—also referred to as echo chambers (Cerf 2016)—that reinforce their beliefs and make them believe that others around the world are more like them (The Wall Street Journal 2016). These echo chambers encourage users to continue passively processing information with a hedonic mindset as they are not confronted with information that runs counter to their beliefs and creates cognitive dissonance. Humans are cognitive misers (Taylor and Fiske 1978), who prefer to process information quickly with a gut-level reaction rather than taking time to consider it critically. Even users in a utilitarian mindset are strongly influenced by their prior beliefs and are likely to process information with a gut-level reaction (Kahneman 2011, Minas et al. 2014), which suggests that the existence of echo chambers on social media may unintentionally keep people processing information without consciously applying a critical filter to what they read.

Humans have two very different cognitive processes for processing information (Kahneman 2011). Many dual process models have been posed under a host of different names; see Evans (2008) for an analysis. In this study, we adopt the commonly used terminology of Keith Stanovich (1999) and Daniel Kahneman (2011), who call these two processes system 1 and system 2. System 1 cognition is automatic and operates continuously, involuntarily providing us with conclusions without our conscious thought (Kahneman 2011). System 1 cognition is our "fast thinking," by which our simple heuristics produce perceptions and actions in less than a second (Kahneman 2011). System 1 is our intuitive decision-making system (Achtziger and Alós-Ferrer 2013); when we have an intuition or a "gut reaction," that is system 1 talking. The quick nature of system 1 is what enables us to

do intuitive tasks without direct thought, such as walk, talk, recognize faces, and effortlessly retrieve certain facts from memory (Kahneman 2011).

However, system 1 comes with certain drawbacks. When we process information using system 1, we only use the information immediately at hand with the vividness and saliency of that available information driving our decisions rather than a more nuanced, carefully considered model (Kahneman 2011, de Castro Bellini-Leite 2013). The associative memory processing of system 1 is strongly influenced by framing (Guo et al. 2017) because it quickly searches for confirming evidence of the question posed. The questions "Is Pat friendly?" and "Is Pat unfriendly?" are fundamentally different questions because they trigger system 1 to retrieve entirely different instances of Pat's behavior (Kahneman 2011). As long as we can form the information into a coherent article—right or wrong—we are likely to follow our immediate system 1 response (Kahneman 2011).

System 2 involves more effortful, deliberate cognition (Kahneman 2011). System 2, or "slow thinking," is laborious and takes more time (Kahneman 2011, Loewenstein et al. 2015). System 2 occurs after system 1—hence, the name system 2 (Kahneman 2011). We have physiological indicators that show the effort involved in system 2 thinking: the pupils dilate, the heart rate changes, the blood pressure rises, and extra blood flows to different areas of the brain that are active (Kahneman 2011). Examples of system 2 processing are doing a computation that requires holding numbers in working memory, monitoring the appropriateness of our behavior in tense social situations, comparing two products, and checking the validity of a complex argument (Mograbi 2011).

Because system 2 cognition requires more effort than system 1 cognition, humans are predisposed to avoid system 2 unless there is a need for it (Stanovich and West 2000, Kahneman 2011). We usually adopt the results of system 1 unless we are motivated to invest effort or system 1 warns us that system 2 should be invoked (Kahneman 2011). One of the most powerful events that can trigger system 2 cognition is a discrepancy from normal expectations (Gersick and Hackman 1990, Louis and Sutton 1991). When system 1 detects that something is not normal, it produces a low "feeling of rightness" (FOR) that indicates something is amiss (Thompson et al. 2011, De Neys 2014). A low FOR is an alert that system 1 believes that system 2 cognition should be invoked for deeper thinking.

Different individuals choose to respond to a low FOR in different ways in different situations. Some individuals have a high need for cognition and, therefore, engage system 2 in situations when the FOR is only slightly low, whereas others may be reluctant to engage in system 2 cognition even when the FOR is very low (Evans and Stanovich 2013, Johnson et al. 2016). Likewise, individuals in a utilitarian mindset who are motivated to make good decisions are more likely to invoke system 2 than individuals in hedonic mindsets who are not motivated to resolve inconsistencies (Kahneman 2011). Empirical evidence suggests that most people's system 1 is relatively good at producing an accurate FOR (Johnson et al. 2016). Whether system 2 is invoked or not depends on the individual's personality and motivation (Johnson et al. 2016). In any event, without a low FOR, social media users are likely to be satisfied with their system 1 response to information they see on social media.

When users encounter information that aligns with their preexisting opinions, their system 1 produces an instant confirmation that the information is sensible, so users are inclined to believe it (Koriat et al. 1980, Devine et al. 1990, Nickerson 1998). When users encounter information that is contrary to their preexisting opinions, their system 1 instantly produces a negative reaction because the information does not align with what system 1 knows to be true, resulting in cognitive dissonance (Festinger 1957). When an individual is presented with two contradictory facts, both of which are plausible (e.g., John is honest, but an article says he lied), the individual must resolve the inconsistency. This can be done either by concluding that the two facts are not contradictory (e.g., John lied, but he is still honest because lying is not related to honesty) or by accepting one and rejecting the other (e.g., John is honest, and thus, I do not believe he lied, or John lied, and thus, I do not believe he is honest) (Festinger 1957).

Resolving such cognitive dissonance takes cognitive effort, and humans often resist invoking system 2 (Taylor and Fiske 1978). This tendency is exacerbated when humans are in a hedonic mindset (Hirschman and Holbrook 1982), but humans still resist contrary information when in a utilitarian mindset (Minas et al. 2014). Because rejecting the new information is easier than reassessing one's preexisting opinions, most people accept their system 1 instant conclusion; they retain their preexisting opinion and discard the new information as being false (Koriat et al. 1980, Devine et al. 1990). Thus, people are more likely to believe information that matches their preexisting opinions (i.e., attitude homophily or alignment; Housholder and LaMarre 2014, Allcott and Gentzkow 2017). This tendency to favor information that confirms one's preexisting opinions and ignore information that challenges them is called confirmation bias (Koriat et al. 1980, Devine et al. 1990, Nickerson 1998). Past research shows that confirmation bias has a

significant effect on the belief in articles posted on social media (Kim et al. 2019, Moravec et al. 2019).

When a user views a news article on social media, the user's system 1 cognition produces an instant assessment of the believability of the article based on its alignment with the user's preexisting opinions on the topic. Social media users are usually in a hedonic mindset, which means they are unlikely to invest the effort needed to invoke system 2 cognition to override their system 1 results. Prior research suggests that seeing quick disclaimers on individual articles (e.g., warnings, reminders, or fake news flags) are often not strong enough to affect belief (Moravec et al. 2019). Although they may trigger cognitive dissonance between the flag and story aligned with preexisting opinions and users may experience low FOR from system 1, it is often not sufficient to change the perceived credibility of the article. The net effect is a belief that is heavily influenced by preexisting opinions.

Online User Ratings

User ratings are used in a variety of contexts, such as product, hotel, restaurant, or service reviews. A confluence of e-commerce research finds that consuming online reviews influences product choice and product purchase (Basuroy et al. 2003, Chevalier and Mayzlin 2006, Hu et al. 2008, Zhu and Zhang 2010, Archak et al. 2011, Lu et al. 2013, Liu and Karahanna 2017). Online reviews can also influence hotel booking intentions and trust (Ye et al. 2009, Sparks and Browning 2011, Banerjee et al. 2017).

However, there is a difference between the *consumption* and the *creation of ratings*. An individual's decision to create a rating can depend on many factors, such as personality (e.g., altruism, narcissism) or a desire to "correct" existing ratings that are perceived to be incorrect (Ho et al. 2017). Prior research suggests that, in utilitarian contexts (e.g., stock market evaluations), we perceive our own ratings to be unbiased (Park et al. 2013) even though we recognize that others' may not be (Ho et al. 2017).

Research on how the act of creating a rating influences our beliefs about that product (or service) is scant, likely because theory argues that our personal experience with the product influences our rating of it. This type of theory does not fit the context of news on social media because very few users have personal experience with the events in the news stories they read. Yet most users quickly take a position (Knobloch-Westerwick and Lavis 2017), often without thinking (Moravec et al. 2019, Pennycook and Rand 2019) because there is no financial incentive to make accurate decisions nor high stakes in what we post on social media. Although some interventions have reduced belief in fake news (Kim et al. 2019, Moravec et al. 2020), research shows that interventions on some posts causes us to believe posts without interventions more (Pennycook et al. 2020).

In summary, past research on online reviews offers little theoretical or empirical guidance on how users rate something with which they have no personal experience. Thus, it remains an open question as to whether the act of creating a rating has an influence on our perceptions of the headlines we read on social media. And, more importantly, are there systematic patterns to this influence that can be used to limit the belief in and spread of fake news?

User Rating of News Articles

Users are commonly asked to rate products and services on the internet. Although some ratings are fake, our focus in this paper is on the act of producing ratings regardless of whether people are accurate in their assessments. When we ask users to rate products or services, we expect them to have used the product or service. When users are prompted to rate, they consider their own experiences and draw upon those experiences to produce the rating. That is, they search their autobiographical episodic memories (i.e., memories associated with events they have personally experienced; Burianova et al. 2010) for experiences related to the use of the product or service and then assess that experience to produce a rating. They may end up choosing a rating that is higher or lower than what they experienced because of a variety of incentives, but that choice does not change their autobiographical episodic memories.

Consider user ratings of news articles: a very small percentage of users have actually been present at the events described in news articles and, thus, have relevant autobiographical episodic memories about the event. For example, hundreds of people experienced the 2018 shooting at Marjory Stoneman Douglas High School in Parkland, Florida. These hundreds of people have relevant autobiographical episodic memories of the shooting, and thus, they are able to comment authoritatively on news stories about it. However, millions of people who were not present during the shooting are also aware of the news article. They have no personal experiences to draw upon, so they have no relevant autobiographical episodic memories about the event and are incapable of providing a direct assessment of the truthfulness of those news articles because they have no personal knowledge of the facts. Therefore, except in very rare circumstances in which the user witnessed the events in the article, users are unable to provide a rating based on personal knowledge. Nonetheless, users are motivated to comment on and share news stories of which they have no personal knowledge (Knobloch-Westerwick and Lavis 2017).

The rarity of personal knowledge of an event in a news article may lead to inaccurate ratings if people feel motivated to provide a rating without having witnessed the events. Some users may search for information and integrate information from various sources, but this does not involve autobiographical episodic memory. In the case of fake news articles that are complete fabrications, there are, of course, no actual events that anyone could have witnessed.

The influence of autobiographical episodic memory retrieval is primarily studied in medical and educational settings (Bluck 2003, Burianova et al. 2010). Autobiographical episodic memory is found to have a beneficial directive function (Bluck 2003) in that our memory of experiences influences behavior. Research also shows that individuals with greater capacity for autobiographical episodic memory exhibit increased conditional reasoning (Barrouillet and Lecas 1999), suggesting a positive relationship between memory and analytical thinking.

We theorize that, when users are asked to rate the truthfulness of an article based on their personal experience, they recognize that they either do or do not have personal knowledge on which to base a rating. This triggers users to face their own level of knowledge on the issue more accurately, which may help them detect fake information. Not all users recognize their lack of personal experience without an appropriate prompt, so we propose a self-referential rating intervention that explicitly asks users if they have personal knowledge of events, thereby encouraging relevant autobiographical memory retrieval or triggering introspection that results in users recognizing that they lack the knowledge needed to rate the truthfulness of the article. Depending on the extent of introspection, users may even devote cognition to process the underlying issue more deeply. When users are asked to provide a rating using a self-referential question that asks about the users' own personal knowledge of the event in the article (e.g., do you have personal knowledge of the event in this article and know the article is true?), most users have no personal knowledge. Thus, their system 1 produces an instant answer that the user has no knowledge and a very low FOR about whether the article is true or not. Of course, for the few users with personal knowledge of the news article, their system 1 provides an answer that they do have knowledge and an answer as to whether the article is true or not with some FOR.

Based on the earlier theorizing, one way to trigger system 2 cognition is by using a discrepancy from normal expectations (Gersick and Hackman 1990, Louis and Sutton 1991). This discrepancy from normal expectations can cause a low FOR, which indicates that something is amiss (Thompson et al. 2011, De Neys 2014). A low FOR is an alert that system 1 believes system 2 cognition should be invoked for deeper thinking. We theorize that when an individual's system 1 is asked whether the user has personal knowledge of the event(s) in the article and returns a negative answer, it produces a low FOR, which is a signal to invoke system 2 cognition. Whether system 2 is invoked or not depends on the individual's personality though a low FOR is sufficient to invoke at least some amount of system 2 cognition under normal conditions (Johnson et al. 2016). The user's system 2 cognition realizes that the user lacks direct knowledge of the events in the article and, thus, cannot produce a truthfulness rating that is based on personal experiences such as with a product or service the user has used. Instead, the user realizes that any initial assessment of truthfulness is based on the credibility of the article's source, the article's fit with the user's preexisting opinions, and indirect knowledge about what is true or false. All of these are weaker criteria than the individual's own direct personal knowledge. The user is forced to reassess the user's knowledge on the topic of the headline and begin thinking analytically. Accordingly, users who invoke their system 2 cognition become less certain of the accuracy of their truthfulness assessment, and they are less likely to believe the news article. As mentioned before, this effect is most pronounced for fake stories, which is when we most need it.

For those few users who do have personal knowledge of an event in a news article, they likely display analytic thinking as a result of what Louis and Sutton (1991) call deliberate initiative. Deliberate initiative is a response to a request for increased conscious attention (e.g., when people are "asked to think" about something) (Louis and Sutton 1991). Rather than critical thinking being triggered by a discrepancy between initial system 1 and system 2 responses, these users are likely triggered to think critically by the explicit prompting to remember events (Louis and Sutton 1991). Greater critical thinking influences the perceived truth of headlines (Pennycook and Rand 2019). We also note that those who actually have personal knowledge of an event are less likely to fall for fake news on that event, so the deliberate initiative serves as an additional layer of defense.

Increased critical thinking increases the activation of regions in the brain that are associated with this type of cognition. The cerebral cortex, which encompasses the upper layers of the brain associated with higher level processing of information, is divided into four distinct lobes: the occipital lobe (central to initial visual processing), the temporal lobe (processing auditory stimuli as well as integrating memories), the parietal lobe (processing and integration of somatosensory experiences), and the frontal cortex (integration of information from lower areas of the brain

as well as integration from the other lobes of the cerebral cortex). The frontal cortex is responsible for the highest level information processing related to consciousness (i.e., system 2 cognition) (Courtney et al. 1998, Williams et al. 2019). Increased activity in the frontal cortex is associated with increased deliberate cognitive activity, including use of working memory and consciousness (Klimesch et al. 1996, 1997, 2001; Krause et al. 2000; Başar et al. 2001; Kilner et al. 2005; Pizzagalli 2007; Moretti et al. 2013).The prefrontal cortex, the most anterior portion of the frontal cortex, is responsible for much of the higher level problem solving, decision making, and information processing in humans (and other primates) (Kaas 2013). Past research finds increased activity in the frontal cortex when users read social media content that interests them (Moravec et al. 2019). Therefore, we propose that the frontal cortex (and specifically the prefrontal cortex) is more active when a person is processing information than when a person is not (Moravec et al. 2019).

In summary, we theorize that we can trigger more critical thinking by changing the design of the current social media interface and that this increase in critical thinking will increase activation of the frontal cortex. Specifically, asking users to evaluate a news article in a way that causes them to consider if they have personal knowledge of the events in the article triggers recognition for most users that they do not have direct personal knowledge and, thus, cannot provide an independent rating (or that they do have direct personal knowledge for those few that experienced the event). This recognition increases the likelihood that users engage in system 2 cognition to assess the article, and the resulting uncertainty reduces the extent to which they believe the article (or change the extent to which they believe the article because of the alignment of their own experience with the way the event is presented in the article). As noted, high-level conscious information processing (i.e., system 2 cognition, whether caused by lack of knowledge or deliberate initiative) is associated with increased activity in the frontal cortex (and its subregions) (Krause et al. 2000, Başar et al. 2001, Pizzagalli 2007, Klimesch 2012, Minas et al. 2018). This leads to two hypotheses:

Hypothesis 1a. *Asking users to rate a news headline using a self-referential question increases cognitive activity in frontal cortical regions.*

Hypothesis 2a. *Asking users to rate a news headline using a self-referential question reduces the extent to which they perceive it to be believable.*

The arguments consider the direct effects of being asked to rate an article on the believability of that specific article, but are there carryover effects? A carryover effect is an effect that occurs after the initial communication (Fischer et al. 2011, Köhler et al. 2017). The communication continues to have an effect in later time periods after the communication is no longer present because it is recalled when users encounter a similar situation in which the communication was presented (Fischer et al. 2011, Köhler et al. 2017). We theorize that, once users realize that they lack direct personal knowledge to rate a specific article, they may generalize this realization to all articles that they have not personally experienced. This realization cannot be "unlearned," and once internalized, it becomes part of the normal assessment of future articles. Thus, once users have some experience with being asked a self-referential question to rate articles (and realize that they lack personal knowledge), they are more likely to exert additional cognitive effort in considering future articles and are less likely to believe those articles (whether asked to rate or not) because they recognize the basis on which one's belief is formed is less certain than a normal assessment.

Moreover, as theorized, our proposed intervention is stronger and provokes increased cognition compared with simple warnings or reminders. This increased cognition carries over to headlines that do not have the intervention as our intervention is not headline-specific. Unlike other interventions that simply tell users whether a specific headline is true or false, our intervention can be equally applied to any headline that a user views because the intervention is not tied directly to the actual veracity of the headline. Given the general nature of our intervention, users who respond with introspection to the initial presentation of the intervention are also more likely to think more deeply about subsequent headlines they encounter without the intervention because subsequent headlines trigger recall of the communication (Fischer et al. 2011, Köhler et al. 2017). Thus, our intervention is likely to have carryover effects. Overall, these effects benefit social media platforms as the use of article ratings without self-referential questions can backfire when users perceive headlines without warnings to be more credible after they observe headlines that have warnings attached (Pennycook et al. 2020).

As noted previously, the prefrontal cortex is central to our reasoning, decision making, and evaluating ourselves in light of our past experiences (Krause et al. 2000, Başar et al. 2001, Pizzagalli 2007, Klimesch 2012, Minas et al. 2018). Studies show that system 2 processing leads to differences in alpha attenuation in the frontal cortex associated with cognitive control, working memory, and attention (Williams et al. 2019). Studies in information systems also show that system 2 cognitive processes such as decision making and confirmation bias (Minas et al. 2014), intuition (Minas et al. 2018), and cognitive dissonance and bias in the

Table 1. Participant Group Description

Facebook use		Political affiliation	
Categories	Percentage	Categories	Percentage
Once a week or less	22	Democrat	31
More than once a week	31	Moderate/independent	15
More than once a day	47	Republican	54

processing of true and false news on social media (Moravec et al. 2019) are associated with differences in alpha attenuation in the frontal cortex. We, therefore, theorize that individuals who have been triggered by a self-referential question process information more deeply. Individuals processing information more deeply show greater activation of their frontal cortex compared with when they are not triggered to think as critically about the accuracy. Thus, we hypothesize the following:

Hypothesis 1b. *There is a carryover effect such that, once users have been asked to rate some headlines using a self-referential question, there is increased cognitive activity in frontal cortical regions for subsequent headlines.*

Hypothesis 2b. *There is a carryover effect such that, once users have been asked to rate some headlines using a self-referential question, they are less likely to perceive subsequent headlines as believable.*

Methodology

Participants

We recruited 68 undergraduate participants from a core business course at a large U.S. university. All were between the ages of 18 and 24, and about 35% of them were female. We also collected Facebook usage and political affiliation information; see Table 1.

Task

The participants viewed 42 news headlines and reported the believability of each article. There were 16 headlines designed to appeal to politically left-leaning participants, 16 to right-leaning participants, and 10 that discussed local events on campus or in the city (see the appendix for the headlines). Half of the headlines were verified as true at the time of the study, and half were verified as false at that time. All headlines were formatted as they might appear as posts on Facebook (see Figure 1).

The headlines and images were designed to avoid major differences in the type and magnitude of feelings they would generate between those headlines appealing to either side of the political spectrum. We used a gender-neutral name for the person posting the news articles—not to be confused with the original source who authored the articles—and the comment from the person posting was a summary of the headline itself. To minimize any news source–specific effect (e.g., some sources are well known and trusted by some users and other sources are not), we fabricated a source name that sounded plausible (*NewsUnion.com*). The URL was verified to be inactive prior to the experiment (i.e., not used by any news provider or anyone else).

Treatments

We used a within-subject, repeated-measures study in which participants received all 42 headlines because this design is required for EEG analysis; between-subjects research designs cannot be used with EEG analyses because brain structures differ between individuals, rendering comparisons meaningless (Moravec et al. 2019).

Figure 1. (Color online) Example News Headline (with the *User-Rating* Treatment)

There were two treatments (control and user-rating) that were identical except for the prompt beneath the post in the user-rating treatment (see Figure 1). Participants responded to the same questions for every headline in both treatments except for the one additional self-referential rating question in the user-rating treatment. For our analysis, we have three separate conditions: control condition (a headline displayed in the control treatment format), user-rating condition (a headline displayed in the experimental treatment format, which asks users for their self-referential rating), and control-after-rating condition (a headline displayed in the same control treatment format that is viewed after seeing one or more prior headlines displayed in the rating format).

In order to test whether the extra cognition is due to the answering of an additional question or to *the question itself causing subjects to realize that they do not personally know whether the article is true*, we test for the difference in cognition between the control condition and the control-after-rating condition (i.e., the carryover effect). We check this increased cognition in both our behavioral and physiological (EEG) analyses.

All participants received both the control condition (no rating) and the experimental treatment (rating condition and control-after-rating condition intermixed). The headlines were randomly assigned to treatments and presented in random order to control for any headline- and image-specific effects. Headlines were randomly assigned to the control-before-rating condition, user-rating condition, and control-after-rating condition and were equally likely to appear in any of the three conditions. Participants saw 14 headlines in the control-before-rating condition. They were then introduced to the rating mechanism. Then, they saw 14 headlines in the user-rating condition and 14 headlines in the control-after-rating condition randomly intermixed. In a repeated-measure design, there is a concern about the effects of an early treatment spilling over into a later treatment (Salkind 2010, Howell 2012). This is usually controlled by random treatment order or a fully crossed design in which all treatment orders are used equally except in cases in which there are likely to be meaningful theoretical differences in the spillover between treatments. This is the case in our study.

The control condition is based on the current Facebook format at the time—thus, unlikely to influence later treatments because it was the interface to which users were accustomed. In contrast, the user-rating condition was likely to have a strong influence on the treatments that followed it because, once users saw a self-referential question, we theorize that they would become more discerning in their consumption of articles; see Hypotheses 1b and 2b. Thus, randomizing the order of the treatments introduces a confound. Therefore, the control condition was always presented first, followed by a description of the user-rating treatment, followed by the experimental treatment with user-rating headlines and control-after-rating headlines intermixed. The overall flow of the experiment is shown in Figure 2. As a robustness check, we also test for any ordering effect that may stem from our experiment design; see our results section.

Participants completed an initial survey for demographic information. The first treatment was the control condition (no user ratings) designed to mimic the current Facebook style of presentation as closely as possible. This treatment had 14 randomly assigned headlines, and users reported the believability of each headline.

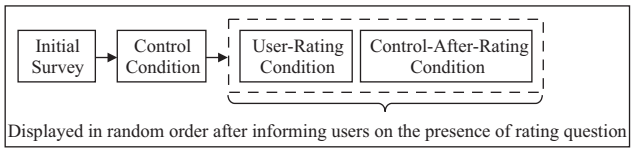
To test for the user-rating effect (Hypotheses 1a and 2a) as well as for the carryover effect (Hypotheses 1b and 2b), the remaining 28 headlines were randomly presented with or without the user-rating treatment. The user-rating treatment presented a self-referential question asking about personal knowledge of the article and whether the article *is* true or *seems* to be true; see Figure 1. Participants answered the self-referential question as they would if it had been implemented on social media, which could result in more cognitive effort simply because they are answering an additional question. However, if this additional cognition is observed in the control-after-rating headlines (for which there is no additional question), then we can conclude that the treatment has effects beyond the increased cognitive effort one would expect from simply answering an additional question.

Independent Variables

We recorded which articles were displayed in the control condition (*control-before-rating*), which were displayed with the user-rating prompt (*user-rating*), and which were displayed without the user-rating prompt but after participants had been exposed to the rating prompt (*control-after-rating*). The control-after-rating headlines appeared exactly as those in the control-before-rating condition.

Confirmation bias was assessed using two items that were self-reported by participants for each headline (Kim et al. 2019). The first was the participant's

Figure 2. Sequence of the Experiment



position on the headline (−3 = extremely negative to 3 = extremely positive). The second was the participant's perceived importance of the headline (using a seven-point scale, 1 = not at all, 7 = extremely). The former represents the *direction* of confirmation bias, whereas the latter represents the *magnitude*. Putting the two together, we are able to measure the degree of fit between a headline and a participant's prior position. The two items are multiplied together to form our *confirmation bias* variable, which ranges from −21 to 21.

Finally, we also controlled for our participants' demographic factors, such as gender, Facebook usage, and political affiliation.

Dependent Variable

The believability of each article was measured by asking each subject to rate each headline (in all three treatments) using three seven-point items (how believable do you find this article, how truthful do you find this article, how credible do you find this article) (Kim et al. 2019). Reliability (Cronbach's alpha) was 0.96, which is adequate.

The other dependent variable in this study is the cognitive activity of participants, measured using EEG. EEG enables the examination of neurophysiological changes that occur during information processing on the order of milliseconds (Berger 1929). EEG measures small electrical signals produced in the superficial areas of the underlying cortical regions. These electrical signals form complex wave patterns at specific frequencies that are related to cognitive activity. We focus on alpha waves because of their potential to indicate memory and consciousness (Pizzagalli 2007).

EEG Analysis

EEG data were collected using an Emotiv EPOC, which is a 14-channel system that is used in prior studies published in leading journals (Minas et al. 2014, 2018; Moravec et al. 2019). In the NeuroIS literature, using consumer-grade systems in analysis of time-frequency data and event-related spectral perturbation (ERSP) analysis has gained broader acceptance as many studies show, for this type of analysis, the consumer-grade systems provide data that is as accurate as larger density, medical-grade systems (albeit with lower spatial resolution from the lower number of electrodes) (see Riedl et al. 2020 for a review). Many studies scrutinize the Emotiv device and find it to obtain a reliable and valid signal of underlying cortical activity—as valid and reliable as larger, high-density systems (Ramírez-Cortes et al. 2010, Debener et al. 2012, Taylor and Schmidt 2012, Badcock et al. 2013, Friedman et al. 2015, Wang et al. 2015). The Emotiv headset has electrodes dispersed over the scalp along the 10-20 system with reference electrodes located at P3 and P4 (Herwig et al. 2003). The system sampled at 128 Hz with a basic finite impulse response high-pass filter of 1 Hz applied to the data. No additional amplifiers are needed. Impedances were verified, and data were collected using Emotiv TestBench Software Version 1.5.0.3.

The data were analyzed using EEGLab, a toolbox using MATLAB (Delorme and Makeig 2004). First, the data are visually inspected to remove eye movement and muscle artifacts. Next, EEGLab's algorithm automatically rejects artifacts greater than three standard deviations from the mean (Delorme and Makeig 2004). EEGLab uses an independent components (IC) analysis (Pizzagalli 2007) to assesses which of the other electrodes account for the most variance in the signal of each individual electrode to produce up to 14 ICs of activation (Onton and Makeig 2006). Each IC represents a pattern of activation over the entire brain, not solely the activity at a specific electrode. Following the procedures of Delorme and Makeig (2004), the data were extracted into epochs (i.e., time windows) and analyzed using the K-means component of EEGLab, which grouped the ICs at the individual participant level into clusters at the experiment level (across all participants) to identify similar patterns of activation. This procedure clusters similar ICs based upon their latency, frequency, amplitude, and scalp distribution with unusual ICs removed into a separate outliers cluster (Onton et al. 2005).

We used a time–frequency analysis technique called ERSP (Makeig 1993). ERSP examines the entire brain over a specified time period to identify brain regions that are active. ERSP is a hypothesis-free technique in which clusters of active brain regions emerge from the data irrespective of any hypotheses the researcher may have. The clusters of regions identified by ERSP may or may not align with the regions the researcher has hypothesized, so when ERSP finds differences in a hypothesized region, it is a powerful signal supporting the theory because nothing directed the analysis to consider that region (Moravec et al. 2019). However, when ERSP analysis finds significant differences in regions about which the researcher did not theorize, the researcher must use abductive reasoning to interpret the results (Moravec et al. 2019).

With ERSP analysis, we look for a pattern of changes in a specific frequency band over a period of several seconds. The time period needs to be consistent for all trials across all participants. The shortest time any participant spent on any headline was just over four seconds, so this set an upper bound for the maximum time period. As is customary, we analyzed the last four seconds the participant viewed the headline so as to focus on the time when the participant was

making the participant's last assessment of the headline before moving on. During this time period, participants were looking at the headline and rating it, which would indicate that, in the treatment group, they were engaging the self-referential question and in a self-referential mindset, whereas in the control group, they exhibit cognition related solely to the act of understanding and reading the headline. Time frames of this duration are often used in previous ERSP studies, including an electronic brainstorming ideation study of alpha activation in the frontal cortex (Minas et al. 2018) and a fake news study examining alpha attenuation during processing of headlines with and without a false news flag (Moravec et al. 2019). We examined the alpha frequency band (8–13 Hz), which is suggested and used in prior research (Klimesch 2012; Minas et al. 2014, 2018; Wang et al. 2014; Müller-Putz et al. 2015; Moravec et al. 2019). Studies of the frontal cortex region suggest that working memory usage may appear more strongly in the upper portion of the alpha band (10–13 Hz) than in the lower portion (8–10 Hz) (Klimesch et al. 1996, 1997, 2001; Kilner et al. 2005; Moretti et al. 2013).

Alpha waves are shown to change reliably in response to stimuli (Klimesch 2012). When a region of the brain becomes active, alpha waves *desynchronize*, leading to lower alpha levels (Cohen 1995); thus, alpha wave desynchronization indicates higher levels of cognitive activity (Makeig et al. 2002, Kelly et al. 2006, Klimesch 2012).

Results

Behavioral Results

To test our hypotheses, we performed hierarchical linear modeling (HLM) (also called mixed modeling, hierarchical modeling, random effects modeling, or random coefficient modeling). We used HLM to mitigate statistical issues with repeated measurement designs by incorporating subject-specific effects as well as within-subject, treatment-level effects (Raudenbush and Bryk 2002, Dennis et al. 2020). Our experimental design is hierarchical in the sense that each subject saw 42 headlines; level 1 of the model is at the headline level, whereas level 2 is at the subject level. The base case at level 1 was the control condition with a true headline, and at level 2, it was politically independent, male, and Facebook use at daily usage (accessing Facebook at least once per day). We used a random intercept model with random slopes, meaning all estimates at level 1 are random. The random slopes also include subject-level covariates (e.g., $\beta_{1j} = \gamma_{10} + \gamma_{1n} subjectcontrols_j + u_{1j}$). The intercept ($\beta_{0j}$) and slope coefficients (β_{1j-4j}) for our headline-level estimates are modeled with a random error term, so the intercept is unique for each subject to account for unhypothesized subject-level effects beyond the effects of gender, Facebook use, and political leaning.

Level 1 (***headline***):

$$Believability_{ij} = \beta_{0j} + \beta_{1j} RatingCondition_{ij} + \beta_{2j} ControlAfterRatingCondition_{ij} + \beta_{3j} FalseHeadline_{ij} + \beta_{4j} ConfirmationBias_{ij} + e_{ij},$$

Level 2 (***subject***):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Female_j + \gamma_{02} FacebookUseOnceaWeek_j + \gamma_{03} FacebookUseLessThanWeekly_j + \gamma_{04} Democrat_j + \gamma_{05} Republican_j + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Female_j + \gamma_{12} FacebookUseOnceaWeek_j + \gamma_{13} FacebookUseLessThanWeekly_j + \gamma_{14} Democrat_j + \gamma_{15} Republican_j + u_{1j},$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Female_j + \gamma_{22} FacebookUseOnceaWeek_j + \gamma_{23} FacebookUseLessThanWeekly_j + \gamma_{24} Democrat_j + \gamma_{25} Republican_j + u_{2j},$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31} Female_j + \gamma_{32} FacebookUseOnceaWeek_j + \gamma_{33} FacebookUseLessThanWeekly_j + \gamma_{34} Democrat_j + \gamma_{35} Republican_j + u_{3j},$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41} Female_j + \gamma_{42} FacebookUseOnceaWeek_j + \gamma_{43} FacebookUseLessThanWeekly_j + \gamma_{44} Democrat_j + \gamma_{45} Republican_j + u_{4j}.$$

The results are shown in Table 2. First, we note that gender, Facebook usage, and political affiliation did not consistently influence believability. Confirmation bias has a significant positive effect, indicating that users are more likely to believe articles that match their preexisting opinions as expected.

Hypothesis 2a posits that users rating articles affects believability. Table 2 shows that user rating has a negative and significant effect on believability, indicating that the user-rating treatment influenced the participants to be more critical about the truthfulness of the headlines they saw. We conclude that Hypothesis 2a is supported.

Hypothesis 2b hypothesizes that user rating would have a carryover effect. Table 2 shows that *control-after-*

Table 2. Estimation Results for Believability

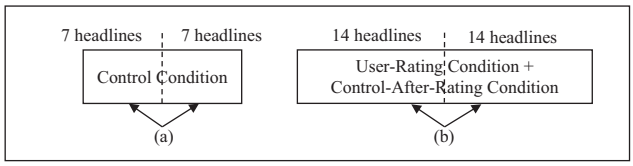
Independent variables	Coefficient
User rating	−0.396**
Female	0.018
FB use: Less than once a week	−0.238
FB use: Less than once a day	0.138
Democrat	0.125
Republican	−0.013
Control-after-rating[a]	−0.550***
Female	−0.086
FB use: Less than once a week	0.210
FB use: Less than once a day	0.563***
Democrat	0.011
Republican	−0.037
False	−1.169***
Female	0.262
FB use: Less than once a week	0.495**
FB use: Less than once a day	0.095
Democrat	0.312
Republican	0.270*
Confirmation bias	0.576***
Female	−0.190*
FB use: Less than once a week	0.322*
FB use: Less than once a day	0.009
Democrat	−0.273**
Republican	−0.040
Model constant	4.438***
Female	0.292
FB use: Less than once a week	−0.269
FB use: Less than once a day	−0.231
Democrat	−0.429
Republican	0.065

Note. Confirmation bias is standardized.
[a]Control-after-rating refers to those headlines displayed without asking users to rate them (i.e., the control condition) after a headline asking users to rate it was presented.
***$p \le 0.001$; **$p \le 0.01$; *$p \le 0.05$.

rating is negative and significant; hence, Hypothesis 2b is also supported. Moreover, the difference between the coefficient for *control-after-rating* and that for *user rating* is not statistically significant ($p = 0.297$), indicating that the carryover effect is as strong as the effect from the rating prompt.

We included one robustness check. As described earlier, we could not adopt a fully crossed experimental design because of the spillover effect; placing the control condition after the user-rating treatment would confound the control treatment (as explained in Hypothesis 2b). The sequence of our experiment (Figure 2) means we cannot rule out ordering effects as an alternative hypothesis for Hypothesis 2b; perhaps users become more skeptical later in the experimental session. To test this, we divided the headlines in each condition into two parts and compared believability between the headlines in the first half of each condition to the headlines in the second half of that respective condition; see Figure 3. We used the same

Figure 3. Testing for Ordering Effect



model as described previously with the exception of headline order replacing treatment condition.

As we can see from Table 3, there was no evidence of any ordering effect; whether the headline was in the first or second half had no effect for the control condition (a) or for the user-rating condition (b). Because there is no evidence to support the rival hypothesis, we conclude that the user-rating treatment influenced users to be more skeptical of the truthfulness of the articles.

Neurophysiological Results

We used ERSP time–frequency analysis to understand the nature of cognition in the control condition (before being asked to rate headlines), in the rating condition (when asked to rate a headline), and in the control-after-rating condition (headlines with no request to rate that were presented after being asked to rate some prior headlines).

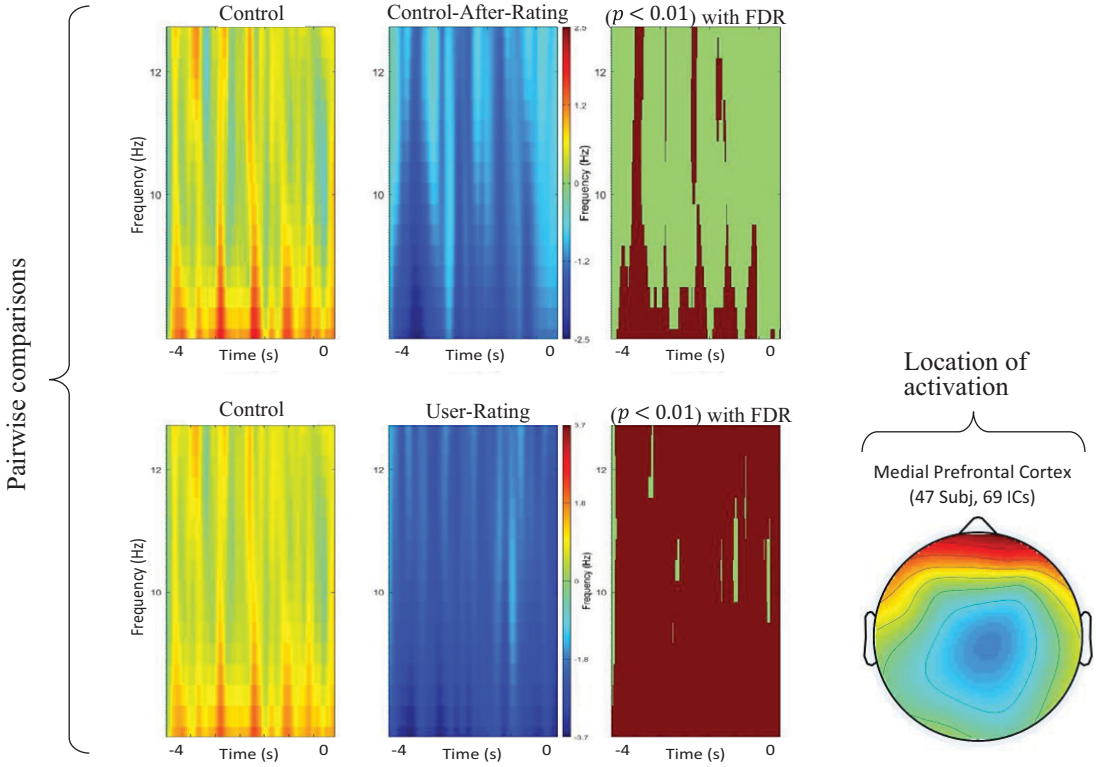
The ERSP analysis identified three clusters of brain regions with activation across our participants. The first cluster shows frontal activation that is slightly right of the midline; see Figure 4. The activation is near the mPFC, which is associated with decision making, memory retrieval and consolidation, and self-reference (Mansouri et al. 2007, 2009; Euston et al. 2012; Kim et al. 2014). The omnibus ERSP test across all three conditions showed statistically significant differences, so we conducted two follow-up pairwise comparisons to test Hypotheses 1a and 1b (shown in Figure 4). The first compares the control condition to the control-after-rating condition (with the false discovery rate (FDR) of Benjamini and Hochberg (1995) to minimize type I error). This finds statistically

Table 3. Estimation Results for Believability

Independent variables	(a) Control	(b) User rating
First half	0.095	0.029
False	−1.068***	−0.804***
Confirmation bias	0.399***	0.515***
Female	0.317	0.209
FB use: More than once a week	−0.311	0.040
FB use: More than once a day	0.036	−0.153
Democrat	−0.195	−0.178
Republican	0.257	0.170
Model constant	4.366***	3.857***

Note. Confirmation bias is standardized.
***$p \le 0.001$; **$p \le 0.01$; *$p \le 0.05$.

Figure 4. (Color online) Differences in Medial Prefrontal Cortex Resulting from Conditions



Notes. The top row shows cognition over time (x-axis) and frequency (y-axis) for headlines in the control condition (left panel) and the control-after-rating condition (middle panel) with darker color indicating more activation. The right panel shows significant differences (in darker gray (print) or red (online)) between the two conditions at $\alpha = 0.01$ with FDR correction. The bottom row shows the same three panels for the control and user-rating conditions. The image on the right shows the location of the activation.

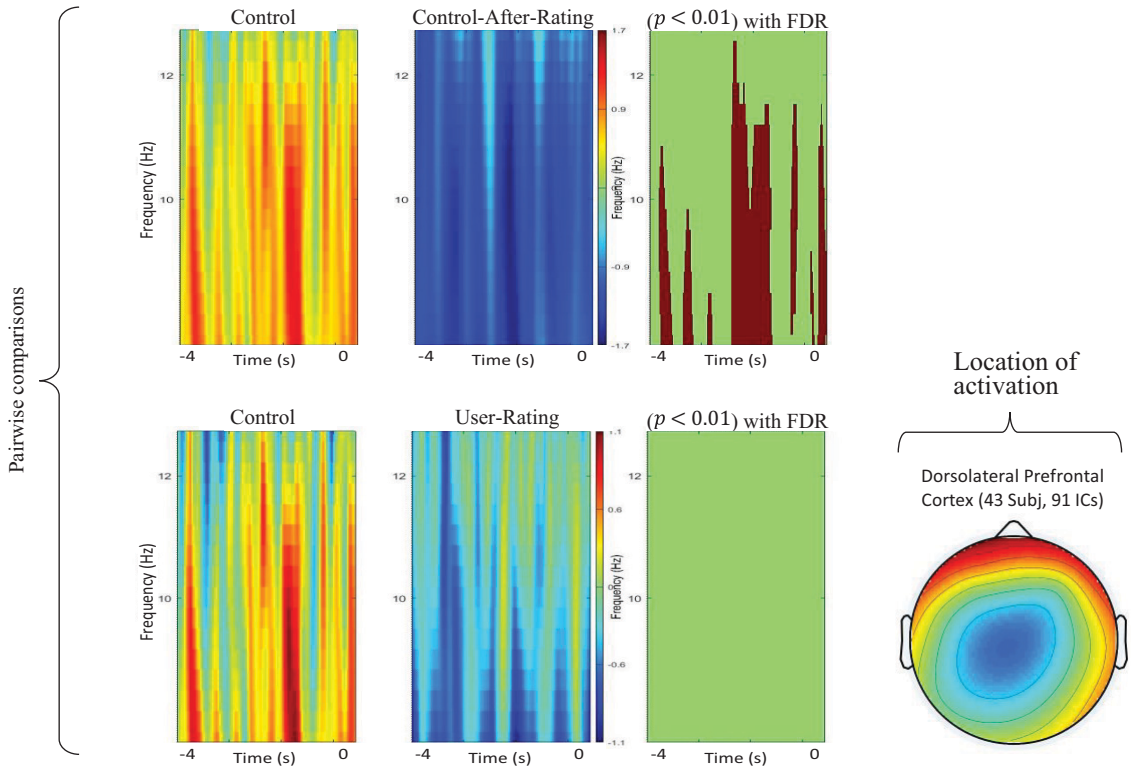
significantly greater activation (i.e., desynchronization) in the control-after-rating condition over some of the alpha band during almost the entire time interval. The second compares the control condition to the rating condition (using FDR) and finds statistically significantly greater activation in the rating condition over almost the entire alpha band during this entire time interval.

This indicates that the rating condition increased activation of the mPFC compared with the control condition, and these changes in cognition persisted in the control-after-rating condition (i.e., the rating condition had a carryover effect when viewing articles that no longer requested ratings). The rating condition is expected to create increased activation in the mPFC because it is closely associated with decision making, retrieving memories, and consolidating new experiences. This pattern occurring in the control-after-rating condition shows that the mPFC process is also occurring after exposure.

The second cluster shows activation on the left of midline; see Figure 5. This area is near the DLPFC, which has many functions. The DLPFC is associated with working memory, executive functioning, and conflict-based behavioral adjustment (Mansouri et al. 2007). The DLPFC is found to be active during conflict resolution by amplifying task-relevant information or suppressing processing of task-irrelevant information (Biswal et al. 1995, Egner and Hirsch 2005, Wang et al. 2014). The omnibus ERSP reveals significant differences among the three treatments across the frequency band. Pairwise comparisons in Figure 5 indicate that the control and ratings conditions do not significantly differ. However, there is significantly more activation (with FDR) in the control-after-rating condition than in the control condition. Prior research shows that the DLPFC is related to behavioral adjustment, which may suggest that the way the individuals approach these articles after having to rate them shifted from how they approached the articles before they were asked to rate them. Alternatively, the activation of the DLPFC may suggest that, after being asked to rate, the participants were amplifying relevant autobiographical information and suppressing irrelevant information.

The third cluster was distributed across the frontal cortex (see Figure 6), and the omnibus test of significance shows significant differences. Pairwise comparisons between the rating and control conditions indicate a few significant differences. There is significantly more activation in the control-after-rating treatment compared with the control treatments mostly in the upper alpha band. This brain region, paired with

Figure 5. (Color online) Differences in Dorsolateral Prefrontal Cortex Resulting from Conditions



findings in the upper alpha band, could indicate that, in the control-after-rating condition, users were experiencing heightened internal awareness as they were exposed to more articles and increased self-reflection (Sauseng et al. 2005, Cahn and Polich 2006, Johnson et al. 2006).

Taken together, the pattern of results in Figures 4–6 shows that the ratings treatment triggered significantly greater activation in different regions of the frontal cortex. The pattern also shows that increased cognition carried over for headlines without an explicit rating request after participants experienced an initial self-referential ratings request. Therefore, we conclude that Hypotheses 1a and 1b are supported.

The presence of the carryover effect—the same pattern observed for control-after-rating as for the rating itself—indicates that this increased cognitive effort is not due to the need to answer an additional question because the control-after-rating condition does not have an additional question. The carryover effect shows that there is increased cognition triggered by subjects assessing whether they had firsthand personal knowledge and occurs regardless of whether the subject was explicitly asked an additional question.

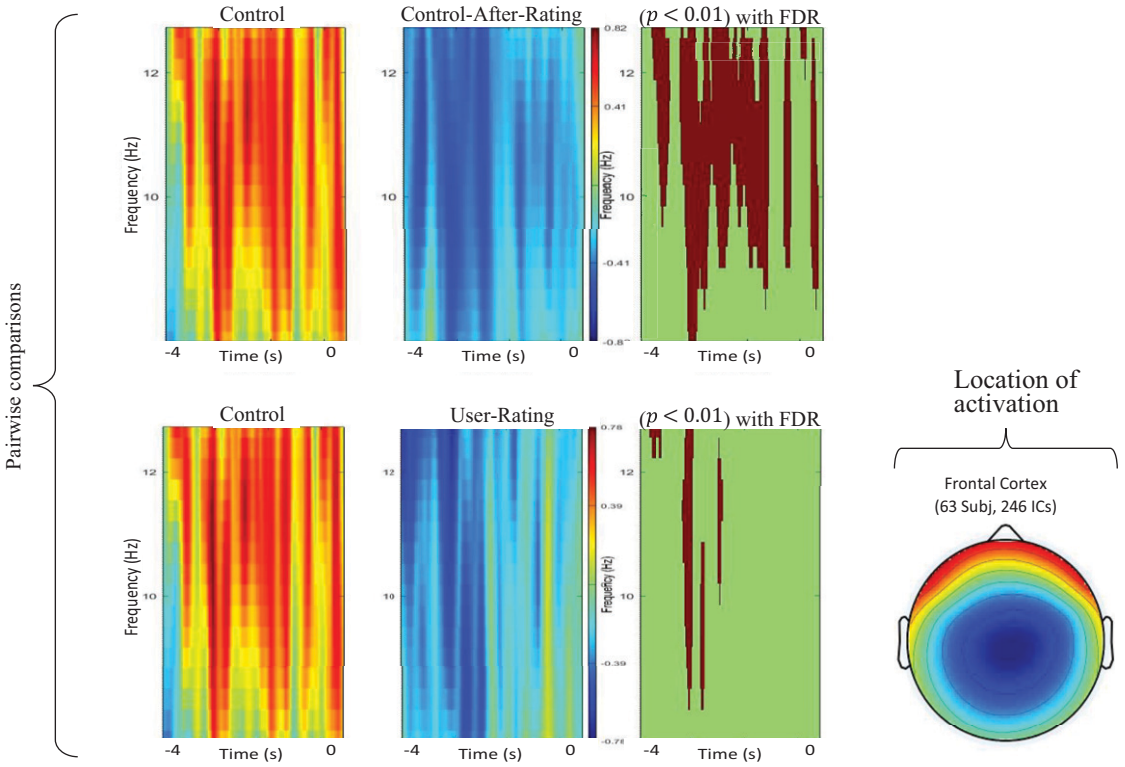
Discussion

Fake news on social media is a major societal problem driven in large part by confirmation bias; users are quick to believe what aligns with their preexisting beliefs and are not motivated to think more deeply about what might be true or false (Moravec et al. 2019). We began this paper by theorizing that crowdsourced user ratings of social media news articles and sources—which suffer from many problems (Dennis et al. 2018)—may provide an as yet untheorized benefit: a self-referential request to rate—which directs users to reflect on their own personal knowledge of the event—may interrupt the usual cycle of confirmation bias.

Understanding how *consuming ratings* affects users is studied in various different contexts, but understanding how the act of *creating a rating* affects beliefs has not been studied—and especially not in social media contexts in which users have no personal knowledge of what they are being asked to rate. A self-referential rating request (one that asks about personal knowledge) may serve as a trigger to think more deeply about an article rather than quickly acceding to one's automatic cognition that immediately suggests articles aligned with prior beliefs are true.

Our results show that confirmation bias remains an important factor influencing beliefs but that prompting a user to rate articles in a way that asks about the user's personal firsthand knowledge made the user less likely to believe articles. This effect persisted for subsequent articles without the rating request. EEG analysis showed that the rating request triggered increased activation in

Figure 6. (Color online) Differences in Frontal Cortex Resulting from Conditions



the mPFC and DLPFC, meaning that there is increased cognition once users are presented with the rating question. As theorized, the rating prompt nudged users to think more critically about the articles by gently asking them whether they *really* knew—from personal experience—if the events described in the articles were true or false. The EEG results support the hypothesis that this question increases cognition relating to memory retrieval, working memory, and self-reflection. Moreover, once this critical filter was activated, it stayed active even for articles without the rating request (control-after-rating headlines). This activation of the critical filter that persisted for control-after-rating headlines demonstrates that the self-referential question triggered additional cognition not because there was an extra question (because the control-after-rating condition had no extra question), but rather that the question induced them to consider their own lack of firsthand knowledge and to continue assessing this even when the self-referential question was not present. Regardless of whether user ratings are a reliable source of rating information, the act of a user self-referentially rating news stories is beneficial for inducing critical thought.

The persistence of asking users to rate articles based on personal experience remedies issues found in previous studies, in which the presence of ratings causes users to believe articles without ratings more (Pennycook et al. 2020). This also highlights the important distinction between the effects of consuming ratings and producing ratings. The impact of the self-referential rating question on perceived believability (see the coefficient for *user rating* in Table 2) was only slightly smaller than the effect of confirmation bias, suggesting that it could be a useful check against misinformation on social media. This intervention moves people away from believing misinformation through self-referential introspection rather than pushing them toward true information. Whereas this intervention is limited in pushing people toward believing true information, we believe the current crisis is embedded in quick belief of misinformation and want to first address this issue before enabling people to believe true information quickly. Social media users become more adept at understanding what is true news once they are better able to detect misinformation accurately.

There are several arguments in favor of user-based ratings besides the cognitive activation described. Developing ratings directly from users may be easier than finding appropriate experts and compensating them for their ratings. There are more users available to rate articles than experts, and voluntary ratings

from users cost nothing. Facebook decided that this is an important step (Dwoskin and Shaban 2018), hopefully adding credibility to news consumption by prioritizing articles from sources that the users have deemed to be reputable. Unfortunately, such crowdsourced rating approaches also have several problems that need to be ironed out before the rating methods can produce a useful reference that can influence users' behavior in a meaningful way (Dennis et al. 2018). After all, while people rate products and services based on their personal experience, the same cannot be expected for news.

However, this very limitation of crowdsourced rating of news—that is, the lack of personal experience of the events described in the news—may have a positive upside. Our work shows that a self-referential question asking users to evaluate the truthfulness of the articles can remind users that they do not have the necessary personal experience to judge whether the articles are true or not, helping them to think more critically about the articles. This effect is important because, as shown in prior research (Kim et al. 2019), believability has strong effects on the actions users take, such as reading, liking, commenting, and sharing. Just as sharing of articles contributes to news—fake or not—going viral, clicking the *like* button or commenting on an article may also spread the news farther and faster because of the appearance of the post in other users' news feeds. Therefore, invoking the critical filter in users' minds—and encouraging them to be more careful in sharing information—is an important first step in tackling fake news.

Implications for Future Research

The prevalence of news consumption on social media (Matsa and Shearer 2018) combined with the hedonic, entertainment-seeking goals of social media use (Sledgianowski and Kulviwat 2009, Johnson and Kaye 2015, Panger 2018) suggest that news consumption on social media may not be mindful. Mindfulness can have significant impacts on the way we use technology for better or worse (Thatcher et al. 2018). In this research, we investigated one approach to inducing more mindful consumption of social media news: consumption that involves invoking deliberate cognition rather than relying on our initial automatic cognition that is heavily influenced by confirmation bias. We believe that asking users to evaluate articles using a self-referential question is one step toward inducing more mindful consumption. Thus, the first implication is that self-referential nudges toward introspection may be successful in increasing user mindfulness. However, we need more research on provoking more mindful social media use, including replications and extensions of this self-referential approach and investigations of other approaches.

Our results show that the creation of user ratings has benefits beyond their consumption, something that has not been deeply considered. Nonetheless, the design of social media supports user ratings. Social media companies have partnered with research institutes to promote surveys asking personal, self-referential questions about products, places, and services (Frauke et al. 2020, Lessler et al. 2021). Hence, our proposed intervention would not be a major deviation from what the users are already used to although using a self-referential question in the specific context of fake news on social media is our novel contribution. Theoretically, we see that small design changes can be sufficiently interesting to users to increase mindfulness. Although the long-term effect remains an open question, the short-term benefits in misinformation detection and rejection may be of more importance currently as we contend with health and political misinformation as well as conspiracy theories.

Thus, a second implication for future research is that the active creation of ratings by users can reduce user belief in misinformation. Future research should assess whether the active creation of ratings can also reduce the spread of misinformation. We found that the presence of a self-referential rating question triggers users to think more critically about the articles presented (for both those with the ratings request and those that follow it); thus, this intervention is well suited to help users disbelieve fake news. We need more research to understand the boundary conditions of this effect and the extent to which it persists for articles without the rating request because our observation was bounded by the length of the experiment. Does it result in permanent changes in cognition and belief, or is it limited to articles in close proximity to the self-referential rating request? We find that participants were thinking more critically a few headlines away, but we need more research to better understand the duration of the carryover effect. If the effect attenuates over time, do repeated requests about first-hand knowledge provide the necessary boost?

Our approach focuses on turning users away from misinformation rather than turning them toward true information. To date, most research on managing misinformation focuses on decreasing belief in misinformation rather than increasing belief in true information. Our focus in this work is to move users away from false information though we recognize that the intervention used in our study may result in a decrease in belief for true news too. We need more research on nudging users toward true information. Nudges to encourage people to believe true information are different in theory than nudges that encourage disbelief in false information as we theorize that humans respond differently to positive and negative information. Often negative information is more surprising

(Yin et al. 2016), which suggests that it may initially be easier to move users away from false information rather than toward true information.

We focus on a self-referential rating request designed to trigger autobiographical episodic memory retrieval as a technique to trigger critical thinking. Some users may actually have firsthand personal knowledge of the event in the headline, but most users do not. By explicitly asking users about their firsthand personal knowledge, we wanted to trigger users to realize that they lacked the knowledge needed to really know if the article was true or false. Such a question was designed to trigger users' automatic cognition (i.e., system 1) to produce a low FOR, which had the intended effect of inducing many users in our study to engage in deliberate cognition (i.e., system 2) as indicated by increased activation in the mPFC and DLPFC. To our knowledge, our study is the first to show a sustained increased pattern of critical thinking triggered by being asked a self-referential question, whereas other studies show that flags cautioning users about headlines may somewhat increase cognition but do not alter belief (Moravec et al. 2019). Thus, we need more research on whether other forms of questions also trigger increased cognition.

For most crowdsourced ratings to be successful, users need to provide accurate ratings, which are then summarized and presented to other users. Our study context is different; the primary benefit is from simply asking users to rate. More research is needed to see whether actually providing the requested rating that would be collected, collated, and posted for others to consume would alter its effects. We theorize it would strengthen the effect as the rating would no longer be hypothetical. Thus, we also need to consider the implications for the consumption of user ratings. Asking users to rate articles means that social media platforms are expected to provide those ratings to other users. Research suggests that, in the context of fake news and social media, users are less accepting of user ratings than ratings produced by experts (Kim et al. 2019) but that research did not consider a rating request that asks about firsthand personal knowledge. For example, some e-commerce websites only present ratings from "verified" purchasers. Ratings that are provided from those with firsthand knowledge may be inherently more trustworthy than more generic ratings.

Conversely, if the presentation of user ratings differentiated between ratings from users with personal knowledge and those without, would it change how user ratings were perceived? The carryover effect that we see in this study from asking about prior knowledge suggests that this information is highly relevant to creating mindfulness. It might also be important for those consuming the ratings, such that users would trust ratings from those with firsthand knowledge more than those from users without firsthand knowledge. The implication is that the same theoretical factors pertaining to personal knowledge at work for the production of user ratings may—or may not—prove to be effective for the consumption of user ratings. We need more research investigating whether it is possible to invoke personal knowledge (and the realization of the lack thereof) during the consumption of ratings, ratings from both users and experts, and if such an invocation has effects on the belief in and spread of fake news.

Another important implication for future research is the role of confirmation bias. Past research shows confirmation bias has a strong effect in driving users to believe fake news (Moravec et al. 2019). From Table 2, we see that the relative strength of asking users to rate articles (a standardized beta coefficient of 0.396) is similar but slightly smaller than the effects of confirmation bias (a beta of 0.576). Here, the request to rate comes close to compensating for confirmation bias, and the difference between the two coefficients is not statistically significant ($p = 0.222$). Thus, a rating question that prompts users to realize that they lack the knowledge to know if the article is really true is a useful tool in reducing confirmation bias. More research is needed on this and interface designs that could be used to trigger this same autobiographical memory process, thereby reducing confirmation bias. For example, instead of a ratings question, could we achieve the same effect by adding a message saying "Were you there? Did you see this?"—the same type of message that news providers sometimes put on breaking stories to solicit eyewitness accounts.

Implications for Practice

The public is recognizing the role of social media and search engine providers in the spread of fake news and is calling for more proactive measures. We approach the problem of fake news from the opposite direction from many other researchers, who have started by first building prototypes and then testing if they affect beliefs (Ratkiewicz et al. 2011, Shao et al. 2016). We started by first testing if a self-referential prompt asking users to rate articles influences them to think more critically about the truthfulness of those articles. In so doing, we aim to provide evidence-based design advice.

Our results show that asking users to rate articles using a self-referential question designed to trigger autobiographic memory retrieval (i.e., do you have personal knowledge of this article?) is likely to induce social media users to engage in greater deliberate cognition about a headline and make them less likely to believe it. The effects are almost as strong as the effects of confirmation bias, suggesting it is a useful tool

in the fight against the belief of and spread of fake news on social media. Social media platforms should implement user ratings, using questions designed to make users realize that they lack personal knowledge of the events and avoid rushed judgments.

Crowdsourced ratings are common in e-commerce but may be challenging for news (Dennis et al. 2018), and expert ratings on news articles may have more impact during the consumption of ratings (Kim et al. 2019). However, rating methods need not be mutually exclusive; we can use different approaches together (e.g., user rating with expert rating). Figuring out the best way to collect and provide rating information must be a long-term endeavor. However, we find that the self-referential user rating has an immediate benefit, and it only requires a small change to the interface.

Conclusion

We investigated a possible positive side effect of asking users to rate news articles on social media. The jury is still out on whether such a crowdsourced rating mechanism would be effective in generating useful information for readers to evaluate the truthfulness of the articles they see. Nonetheless, our work shows that there is value in asking users to rate news articles using a self-referential question regardless of whether those ratings are ever used by others because it triggers users to realize that they may not know if the articles are really true.

Appendix. News Headlines Used in the Experiment

1	Robert Mueller Encouraged to Resign over Affair with Nancy Pelosi
2	Body of Clintons' Former Housekeeper Found in Arkansas
3	Nancy Pelosi Said Building a Wall Will Violate Rights of "Millions of Illegals"
4	Senator Tom Cotton Called for Drug Testing for Social Security Recipients
5	New Law Allows Atheist Doctors to Refuse Care to Religious Patients
6	Democrats "Up All Night Doing Blow" with Obama to Celebrate Shutdown
7	Hillary Clinton Gave 20% of U.S. Uranium to Russia in Exchange for Clinton Foundation Donations
8	Texas Church Shooter Was an Atheist on the DNC Payroll
9	Women's March Speaker Donna Hylton Served Time for Murdering a Man
10	White House Changed Outgoing Message on Comment Line to Blame Democrats
11	DNC Staffer Faces Scrutiny Over "Cisgender Straight White Males" Email
12	Trump Calls for Death Penalty for NYC Truck Attack Suspect
13	DNC and Clinton Campaign Funded the Trump/Russia "Steele Dossier"
14	EPA Reportedly Eliminates Program Helping Cities Deal with Climate Change
15	Trump Administration Declines to Renew EPA's Science Advisor Appointments
16	New York Lawmaker Negotiating Anti-Harassment Laws Is Accused of Sexual Misconduct
17	NFL Players Union Votes to Encourage Kneeling During National Anthem
18	Melania Trump Bans White House Staff from Taking Flu Shot
19	Trump's Doctor Said the President Was "Too Sick to Talk to Mueller"
20	Newsweek Reports That Trump Will Be Impeached and Replaced by Hillary Clinton
21	Jared Kushner Traveled to Saudi Arabia Because It Doesn't Have an Extradition Treaty with the U.S.
22	Iceland Mandates Mental Health Warnings on All Bibles
23	Virginia'' Governor Legalizes Marijuana Across Entire State
24	John McCain Caused a Fire Aboard the USS Forrestal That Killed 134 People
25	Donald Trump Sold an Apartment to a Brutal Haitian Dictator in 1983
26	Kochs Contribute $500,000 to Paul Ryan After the GOP Tax Plan Was Passed
27	United States Senate Candidate Favors a Woman's Right to Cook Him Dinner
28	Trump Overstates Missing Texts Involving FBI Agent
29	Sen. Tom Cotton Issued a "Cease and Desist" Letter to a Constituent
30	President Trump Plans to "End" a Program Funding Heating for the Elderly and Disabled
31	White House Releases Report Contradicting Its Own Position on Climate Change
32	Anti-Abortion Congressman Tim Murphy Urges His Mistress to Have an Abortion
33	[State University] Forms Committee to Investigate Designated Smoking Areas in Dorms[a]
34	[State University] Expanding Code of Conduct to Prohibit Hate Speech[a]
35	[State] Senate Passes Bill Requiring Schools to Get Parent Consent Before Teaching Sex Ed[a]
36	[State] Senate Decided Not to Hear Hate Crime Bill Due to Disagreements in Language[a]
37	Bill to Increase Smoking Age to 18 Passed [State] House, but Died Due to Fiscal Concerns[a]
38	[Local state] Lawmakers Plan to Allow Church Hour Sunday Alcohol Sales[a]
39	New [Local] County Food and Beverage Tax Only Incurred by Out of State Residents[a]
40	[State] Senate Passes Bill Legalizing Cannabidiol Oil[a]
41	[State] House Working on Bill to Require Women Undergoing Abortions to Sign Instructions[a]
42	[Local] County Plans to Make Bid to Be First [State] County to Legalize Weed[a]

[a]Local headlines blinded for review purposes.

References

Achtziger A, Alós-Ferrer C (2013) Fast or rational? A response-times study of Bayesian updating. *Management Sci.* 60(4):923–938.

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. *J. Econom. Perspect.* 31(2):211–236.

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* 57(8):1485–1509.

Badcock NA, Mousikou P, Mahajan Y, de Lissa P, Thie J, McArthur G (2013) Validation of the EMOTIV EPOC® EEG gaming system for measuring research quality auditory ERPS. *PeerJ* 1:e38.

Banerjee S, Bhattacharyya S, Bose I (2017) Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems* 96:17–26.

Barrouillet P, Lecas J-F (1999) Mental models in conditional reasoning and working memory. *Thinking Reasoning* 5(4):289–302.

Barthel M, Mitchell A, Holcomb J (2016) Many Americans believe fake news is sowing confusion, Pew Research Center (15), Washington, D.C., https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.

Başar E, Başar-Eroglu C, Karakaş S, Schürmann M (2001) Gamma, alpha, delta, and theta oscillations govern cognitive processes. *Internat. J. Psychophysiology* 39(2–3):241–248.

Basuroy S, Chatterjee S, Ravid SA (2003) How critical are critical reviews? The box office effects of film critics, star power, and budgets. *J. Marketing* 67(4):103–117.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B* 57(1):289–300.

Berger H (1929) On the EEG in humans. *Arch Psychiatr Nervenkr* 87: 527–570.

Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance Medicine* 34(4):537–541.

Bluck S (2003) Autobiographical memory: Exploring its functions in everyday life. *Memory* 11(2):113–123.

Burianova H, McIntosh AR, Grady CL (2010) A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *Neuroimage* 49(1):865–874.

Cahn BR, Polich J (2006) Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psych. Bull.* 132(2):180–211.

Cerf VG (2016) Information and misinformation on the internet. *Comm. ACM* 60(1):9.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Cohen L (1995) *Time-Frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ).

Cotte J, Chowdhury TG, Ratneshwar S, Ricci LM (2006) Pleasure or utility? Time planning style and web usage behaviors. *J. Interactive Marketing* 20(1):45–57.

Courtney S, Petit L, Haxby J, Ungerleider L (1998) The role of prefrontal cortex in working memory: Examining the contents of consciousness. *Philos. Trans. Roy. Soc. London Ser. B.* 353(1377): 1819–1828.

Debener S, Minow F, Emkes R, Gandras K, Vos M (2012) How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology* 49(11):1617–1621.

de Castro Bellini-Leite S (2013) The embodied embedded character of system 1 processing. *Mens Sana Monographs* 11(1):239.

Delorme A, Makeig S (2004) EEGlab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neuroscience Methods* 134(1):9–21.

De Neys W (2014) Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking Reasoning* 20(2):169–187.

Dennis AR, Kim A, Moravec PL (2018) Facebook's bad idea: Crowdsourced ratings work for toasters, but not news. Accessed January 20, 2018, https://www.buzzfeed.com/alandennis/facebooks-bad-idea-crowsourced-ratings-work-for-toasters.

Dennis AR, Yuan L, Feng X, Webb E, Hsieh CJ (2020) Digital nudging: Numeric and semantic priming in e-commerce. *J. Management Inform. Systems* 37(1):39–65.

Devine PG, Hirt ER, Gehrke EM (1990) Diagnostic and confirmation strategies in trait hypothesis testing. *J. Personality Soc. Psych.* 58 (6):952–963.

Dwoskin E, Shaban H (2018) Facebook will now ask users to rank news organizations they trust. *The Washington Post.* Accessed January 20, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/01/19/facebook-will-now-ask-its-users-to-rank-news-organizations-they-trust/.

Egner T, Hirsch J (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience* 8(12):1784–1790.

Euston DR, Gruber AJ, McNaughton BL (2012) The role of medial prefrontal cortex in memory and decision making. *Neuron* 76(6):1057–1070.

Evans JSBT (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Rev. Psych.* 59(1):255–278.

Evans JSBT, Stanovich KE (2013) Dual-process theories of higher cognition: Advancing the debate. *Perspect. Psych. Sci.* 8(3):223–241.

Festinger L (1957) *A Theory of Cognitive Dissonance Evanston* (Row, Peterson, IL).

Fischer M, Albers S, Wagner N, Frie M (2011) Dynamic marketing budget allocation across countries, products, and marketing activities. *Marketing Sci.* 30(4):568–585.

Frauke K, Neta B, Alyssa B, Adrianne B, Samantha C, Roee E, Junchuan F, et al. (2020) Partnering with a global platform to inform research and public policy making. *Surveys Res. Methods* 14(2):159–163.

Friedman D, Shapira S, Jacobson L, Gruberger M (2015) A data-driven validation of frontal EEG asymmetry using a consumer device. *Internat. Conf. Affective Comput. Intelligent Interaction*, 930–937.

Gabielkov M, Ramachandran A, Chaintreau A, Legout A (2016) Social clicks: What and who gets read on Twitter? *ACM SIGMETRICS/IFIP Performance 2016*.

Gersick CJG, Hackman JR (1990) Habitual routines in task-performing groups. *Organ. Behav. Human Decision Processes* 47(1):65–97.

Graves L (2016) Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Stud.* 19(5):613–631.

Guo L, Trueblood JS, Diederich A (2017) Thinking fast increases framing effects in risky decision making. *Psych. Sci.* 28(4): 530–543.

Herwig U, Satrapi P, Schönfeldt-Lecuona C (2003) Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topography* 16(2):95–99.

Hirschman EC, Holbrook MB (1982) Hedonic consumption: Emerging concepts, methods and propositions. *J. Marketing* 46(3): 92–101.

Ho Y-C, Wu J, Tan Y (2017) Disconfirmation effect on online rating behavior: A structural model. *Inform. Systems Res.* 28(3):626–642.

Housholder EE, LaMarre HL (2014) Facebook politics: Toward a process model for achieving political source credibility through social media. *J. Inform. Tech. Politics* 11(4):368–382.

Howell DC (2012) *Statistical Methods for Psychology* (Cengage Learning, Boston).

Hu N, Liu L, Zhang JJ (2008) Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Inform. Tech. Management* 9(3):201–214.

Johnson TJ, Kaye BK (2015) Reasons to believe: Influence of credibility on motivations for using social networks. *Comput. Human Behav.* 50:544–555.

Johnson ED, Tubau E, De Neys W (2016) The doubting system 1: Evidence for automatic substitution sensitivity. *Acta Psychologica* 164:56–64.

Johnson MK, Raye CL, Mitchell KJ, Touryan SR, Greene EJ, Nolen-Hoeksema S (2006) Dissociating medial frontal and posterior cingulate activity during self-reflection. *Soc. Cognitive Affective Neuroscience* 1(1):56–64.

Kaas JH (2013) The evolution of brains from early mammals to humans. *Wiley Interdisciplinary Rev. Cognitive Sci.* 4(1):33–45.

Kahneman D (2011) *Thinking, Fast and Slow* (Farrar, Straus, and Giroux, New York).

Kelly SP, Lalor EC, Reilly RB, Foxe JJ (2006) Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *J. Neurophysiology* 95(6):3844–3851.

Kilner JM, Mattout J, Henson R, Friston KJ (2005) Hemodynamic correlates of EEG: A heuristic. *Neuroimage* 28(1):280–286.

Kim A, Dennis AR (2019) Says who? The effects of presentation format and source rating on fake news in social media. *Management Inform. Systems Quart.* 43(3):1025–1039.

Kim C, Johnson NF, Gold BT (2014) Conflict adaptation in prefrontal cortex: Now you see it, now you don't. *Cortex* 50:76–85.

Kim A, Moravec PL, Dennis AR (2019) Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *J. Management Inform. Systems* 36(3):931–968.

Klimesch W (2012) Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cognitive Sci.* 16(12):606–617.

Klimesch W, Doppelmayr M, Pachinger T, Ripper B (1997) Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neuroscience Lett.* 238(1):9–12.

Klimesch W, Doppelmayr M, Stadler W, Pöllhuber D, Sauseng P, Röhm D (2001) Episodic retrieval is reflected by a process specific increase in human electroencephalographic theta activity. *Neuroscience Lett.* 302(1):49–52.

Klimesch W, Schimke H, Doppelmayr M, Ripper B, Schwaiger J, Pfurtscheller G (1996) Event-related desynchronization (ERD) and the DM effect: Does alpha desynchronization during encoding predict later recall performance? *Internat. J. Psychophysiology* 24(1–2):47–60.

Knobloch-Westerwick S, Lavis SM (2017) Selecting serious or satirical, supporting or stirring news? Selective exposure to partisan versus mockery news online videos. *J. Comm.* 67(1):54–81.

Köhler C, Mantrala MK, Albers S, Kanuri VK (2017) A meta-analysis of marketing communication carryover effects. *J. Marketing Res.* 54(6):990–1008.

Koriat A, Lichtenstein S, Fischhoff B (1980) Reasons for confidence. *J. Experiment. Psych. Human Learn.* 6(2):107–118.

Krause CM, Sillanmäki L, Koivisto M, Saarela C, Häggqvist A, Laine M, Hämäläinen H (2000) The effects of memory load on event-related EEG desynchronization and synchronization. *Clinical Neurophysiology* 111(11):2071–2078.

Lessler J, Grabowski MK, Grantz KH, Badillo-Goicoechea E, Metcalf CJE, Lupton-Smith C, Azman AS, Stuart EA (2021) Household covid-19 risk and in-person schooling. *Sci.* 372 (6546):1092–1097.

Liu QB, Karahanna E (2017) The dark side of reviews: The swaying effects of online product reviews on attribute preference construction. *Management Inform. Systems Quart.* 41(2):427–448.

Loewenstein G, O'Donoghue T, Bhatia S (2015) Modeling the interplay between affect and deliberation. *Decision* 2(2):55–81.

Louis MR, Sutton RI (1991) Switching cognitive gears: From habits of mind to active thinking. *Human Relations* 44(1):55–76.

Lowrey W (2017) The emergence and development of news fact-checking sites: Institutional logics and population ecology. *Journalism Stud.* 18(3):376–394.

Lu X, Ba S, Huang L, Feng Y (2013) Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Inform. Systems Res.* 24(3):596–612.

Makeig S (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography Clinical Neurophysiology* 86(4):283–293.

Makeig S, Westerfield M, Jung T-P, Enghoff S, Townsend J, Courchesne E, Sejnowski T (2002) Dynamic brain sources of visual evoked responses. *Sci.* 295(5555):690–694.

Mansouri FA, Buckley MJ, Tanaka K (2007) Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. *Sci.* 318(5852):987–990.

Mansouri FA, Tanaka K, Buckley MJ (2009) Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Rev. Neuroscience* 10(2):141–152.

Matsa KE, Shearer E (2018) News use across social medial platforms 2018. Pew Research Center, Washington, D.C. Accessed November 5, 2019, https://www.pewresearch.org/journalism/2018/09/10/news-use-across-social-media-platforms-2018/.

Meixler E (2017) Facebook is dropping its fake news red flag warning after finding it had the opposite effect. *Time.* Accessed June 13, 2018, https://time.com/5077002/facebook-fake-news-articles/.

Minas RK, Dennis AR, Potter RF, Kamhawi R (2018) Triggering insight: Using neuroscience to understand how priming changes individual cognition during electronic brainstorming. *Decision Sci.* 49(5):788–826.

Minas RK, Potter RF, Dennis AR, Bartelt V, Bae S (2014) Putting on the thinking cap: Using neuroIS to understand information processing biases in virtual teams. *J. Management Inform. Systems* 30(4):49–82.

Mograbi GJC (2011) Neural basis of decision-making and assessment: Issues on testability and philosophical relevance. *Mens Sana Monographs* 9(1):251–259.

Moravec PL, Kim A, Dennis A (2020) Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Inform. Systems Res.* 31(3):987–1006.

Moravec PL, Minas RK, Dennis AR (2019) Fake news on social media: People believe what they want to believe when it makes no sense at all. *Management Inform. Systems Quart.* 43(4):1343–1360.

Moretti DV, Paternicò D, Binetti G, Zanetti O, Frisoni GB (2013) EEG upper/low alpha frequency power ratio relates to temporo-parietal brain atrophy and memory performances in mild cognitive impairment. *Frontiers Aging Neuroscience* 5:63.

Mosseri A (2016) Addressing hoaxes and fake news. Facebook Newsroom: Facebook. Accessed February 13, 2020, https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/.

Müller-Putz GR, Riedl R, Wriessnegger SC (2015) Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Comm. Assoc. Inform. Systems* 37:46.

Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. General Psych.* 2(2):175–220.

Onton J, Makeig S (2006) Information-based modeling of event-related brain dynamics. Neuper C, Klimesch W, eds. *Progress in Brain Research*, vol. 159 (Elsevier, Amsterdam), 99–120.

Onton J, Delorme A, Makeig S (2005) Frontal midline EEG dynamics during working memory. *Neuroimage* 27(2):341–356.

Panger G (2018) People tend to wind down, not up, when they browse social media. *Proc. ACM Human-Comput. Interaction*, 1–29.

Park J, Konana P, Gu B, Kumar A, Raghunathan R (2013) Information valuation and confirmation bias in virtual communities: Evidence from stock message boards. *Inform. Systems Res.* 24(4): 1050–1067.

Pennycook G, Rand DG (2019) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188:39–50.

Pennycook G, Bear A, Collins ET, Rand DG (2020) The implied truth effect: Attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. *Management Sci.* 66(11):4944–4957.

Pizzagalli DA (2007) Electroencephalography and high-density electrophysiological source localization. Cacioppo JT, Tassinary LG, Berntson GG, eds. *Handbook of Psychophysiology* (Cambridge University Press), 56–84.

Ramírez-Cortes JM, Alarcon-Aquino V, Rosas-Cholula G, Gomez-Gil P, Escamilla-Ambrosio J (2010) P-300 rhythm detection using ANFIS algorithm and wavelet feature extraction in EEG signals. *Proc. World Congress Engrg. Comput. Sci.* (International Association of Engineers, San Francisco), 963–968.

Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F (2011) Truthy: mapping the spread of astroturf in microblog streams. *Proc. 20th Internat. Conf. Companion World Wide Web*.

Raudenbush SW, Bryk AS (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage Publications, Inc., Thousand Oaks, CA).

Riedl R, Minas RK, Dennis AR, Müller-Putz GR (2020) Consumer-grade EEG instruments: Insights on the measurement quality based on a literature review and implications for neuroIS research. *NeuroIS Retreat* (Springer), 350–361.

Salkind NJ (2010) *Encyclopedia of Research Design* (Sage Publications, Inc., Thousand Oaks, CA).

Sauseng P, Klimesch W, Schabus M, Doppelmayr M (2005) Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *Internat. J. Psychophysiology* 57(2):97–103.

Shao C, Ciampaglia GL, Flammini A, Menczer F (2016) Hoaxy: A platform for tracking online misinformation. *Proc. 25th Internat. Conf. Companion World Wide Web*.

Silverman C (2016) This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed news. Accessed January 25, 2017, https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

Sledgianowski D, Kulviwat S (2009) Using social network sites: The effects of playfulness, critical mass and trust in a hedonic context. *J. Comput. Inform. Systems* 49(4):74–83.

Sparks BA, Browning V (2011) The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management* 32(6):1310–1323.

Stanovich KE (1999) *Who Is Rational? Studies of Individual Differences in Reasoning* (Psychology Press).

Stanovich KE, West RF (2000) Individual differences in reasoning: Implications for the rationality debate? *Behav. Brain Sci.* 23(5): 645–665.

Taylor SE, Fiske ST (1978) Salience, attention, and attribution: Top of the head phenomena. Berkowitz L, ed. *Advances in Experimental Social Psychology* (Academic Press), 249–288.

Taylor GS, Schmidt C (2012) Empirical evaluation of the EMOTIV EPOC BCI headset for the detection of mental actions. Proc. Human Factors Ergonomics Soc. Annual Meeting, vol. 56, 193–197.

Thatcher JB, Wright RT, Sun H, Zagenczyk TJ, Klein R (2018) Mindfulness in information technology use: Definitions, distinctions, and a new measure. *Management Inform. Systems Quart.* 42(3): 831–847.

The Wall Street Journal (2016) Blue feed, red feed. *The Wall Street Journal Online*, accessed February 1, 2017, http://graphics.wsj.com/blue-feed-red-feed/.

Thompson VA, Prowse Turner JA, Pennycook G (2011) Intuition, reason, and metacognition. *Cognitive Psych.* 63(3):107–140.

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Sci.* 359(6380):1146–1151.

Wang S, Gwizdka J, Chaovalitwongse WA (2015) Using wireless EEG signals to assess memory workload in the *N*-back task. *IEEE Trans. Human-Machine Systems* 46(3):424–435.

Wang T, Chen Z, Zhao G, Hitchman G, Liu C, Zhao X, Liu Y, Chen A (2014) Linking inter-individual differences in the conflict adaptation effect to spontaneous brain activity. *Neuroimage* 90: 146–152.

Williams CC, Kappen M, Hassall CD, Wright B, Krigolson OE (2019) Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *Neuroimage* 189:574–580.

Wintersieck AL (2017) Debating the truth: The impact of fact-checking during electoral debates. *Amer. Political Res.* 45(2):304–331.

Ye Q, Law R, Gu B (2009) The impact of online user reviews on hotel room sales. *Internat. J. Hospitality Management* 28(1): 180–182.

Yin D, Mitra S, Zhang H (2016) Research note—When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Inform. Systems Res.* 27(1):131–144.

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133–148.